NPACI– IPG Collaborators Program-Level Overview

George Kremenek
kremenek@sdsc.edu

San Diego Supercomputer Center,
NPACI UCSD

NPACI Technology Collaborations

- **Data management systems**
- **Grid Services Support**
- **Grid portals**
- **Parameter sweep scheduling**
- **Network Weather Service**

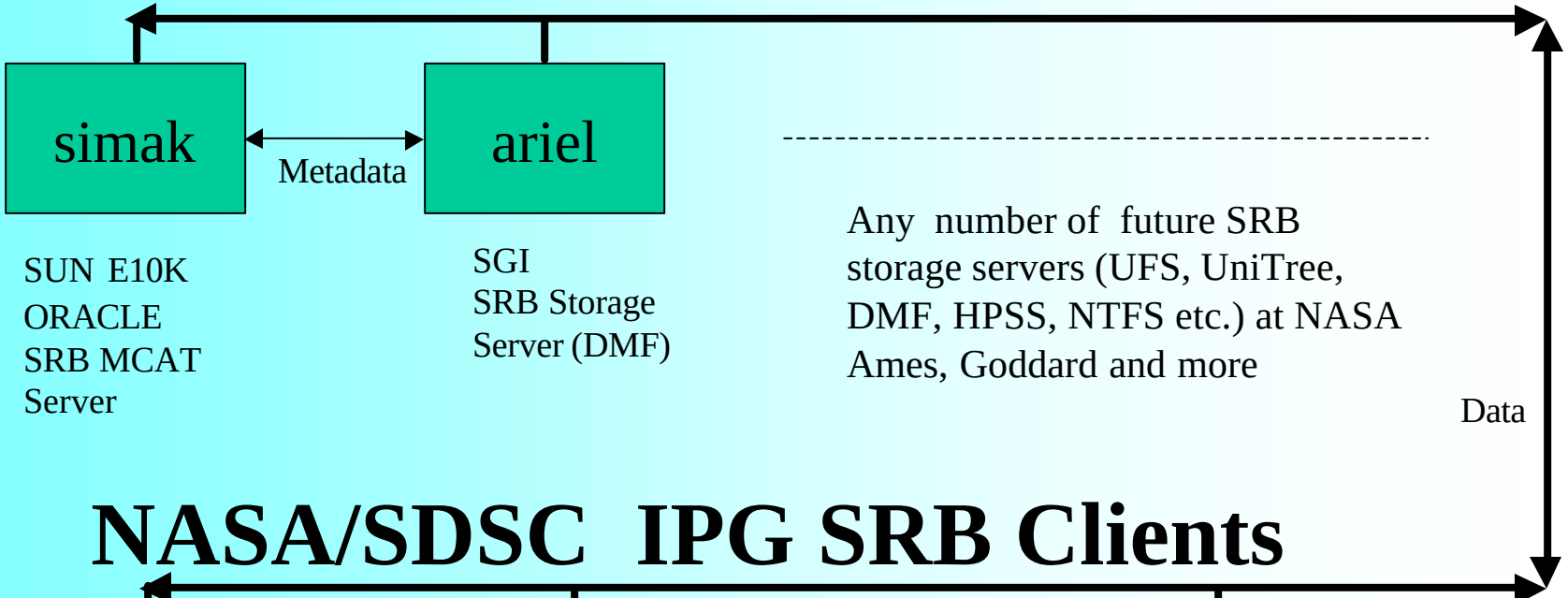
NPACI Supported Projects

- **GridFTP/SRB interface**
- **SRB Collection Management System Installation at NASA Ames**
- **Telescience Milestone**
- **Distributed data processing, parallel transfers, 50+ Mbits/sec**

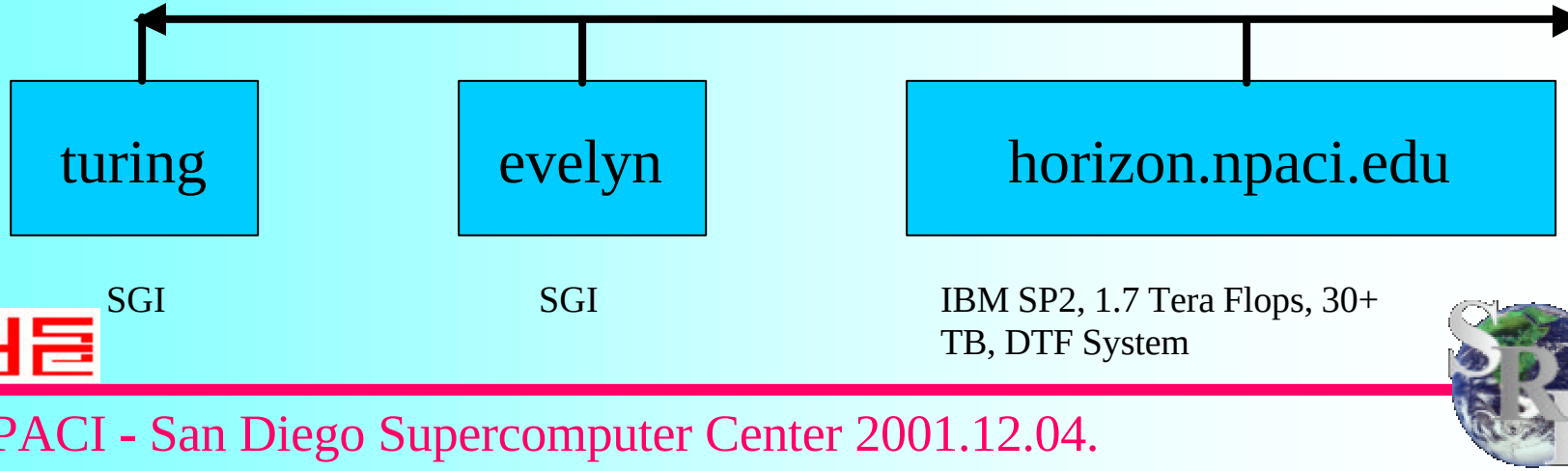
SRB at NASA Systems

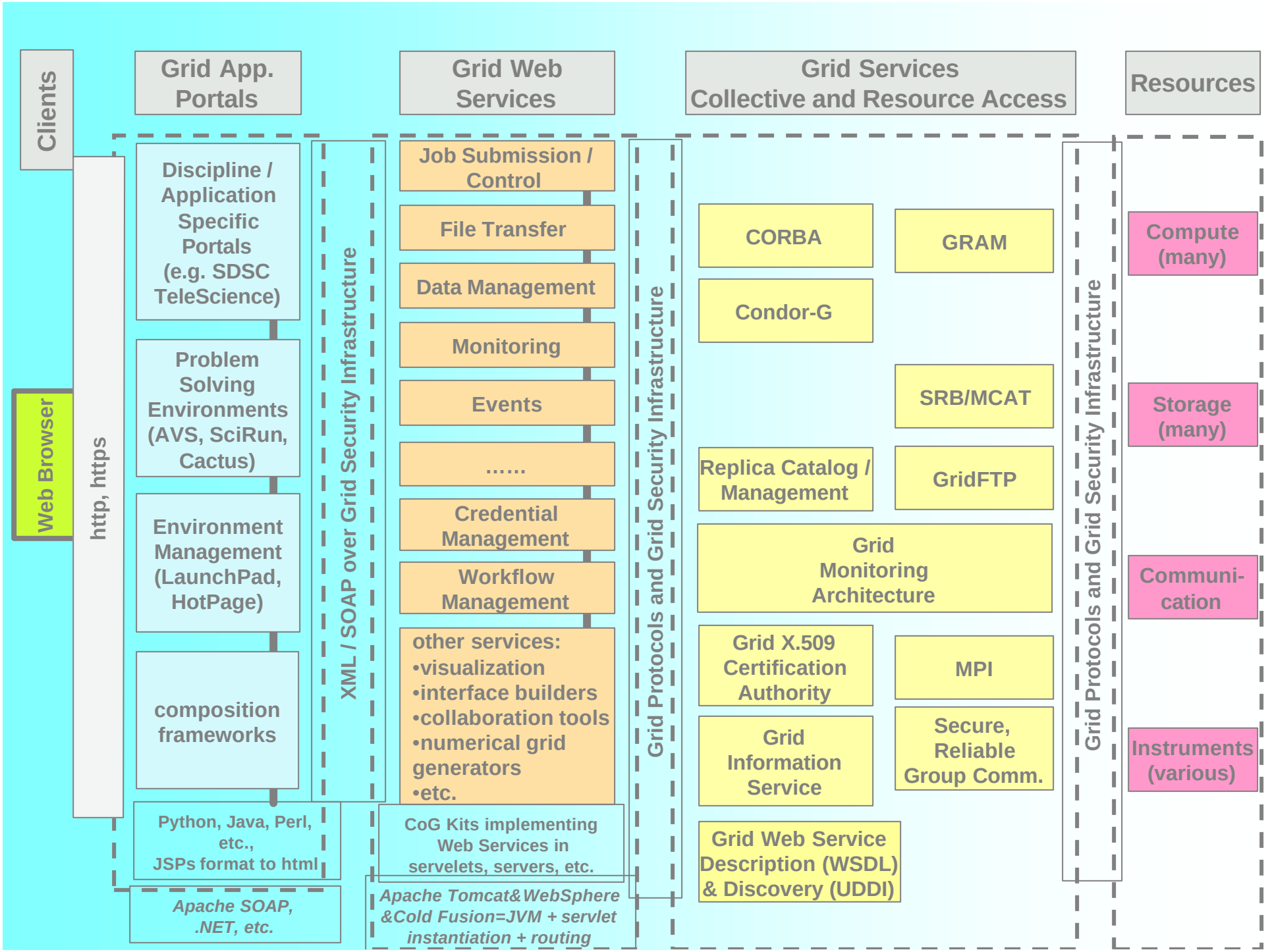
- SDSC - SRB/MCAT (srb@sdsc.edu)
  - Federated collection management system
  - Logical collections than spans multiple storage systems and storage protocols
  - Location transparency, storage protocol transparency, replication, containers
  - Open and persistent storage technology
  - Supports GSI authentication

NASA IPG SRB Servers



NASA/SDSC IPG SRB Clients



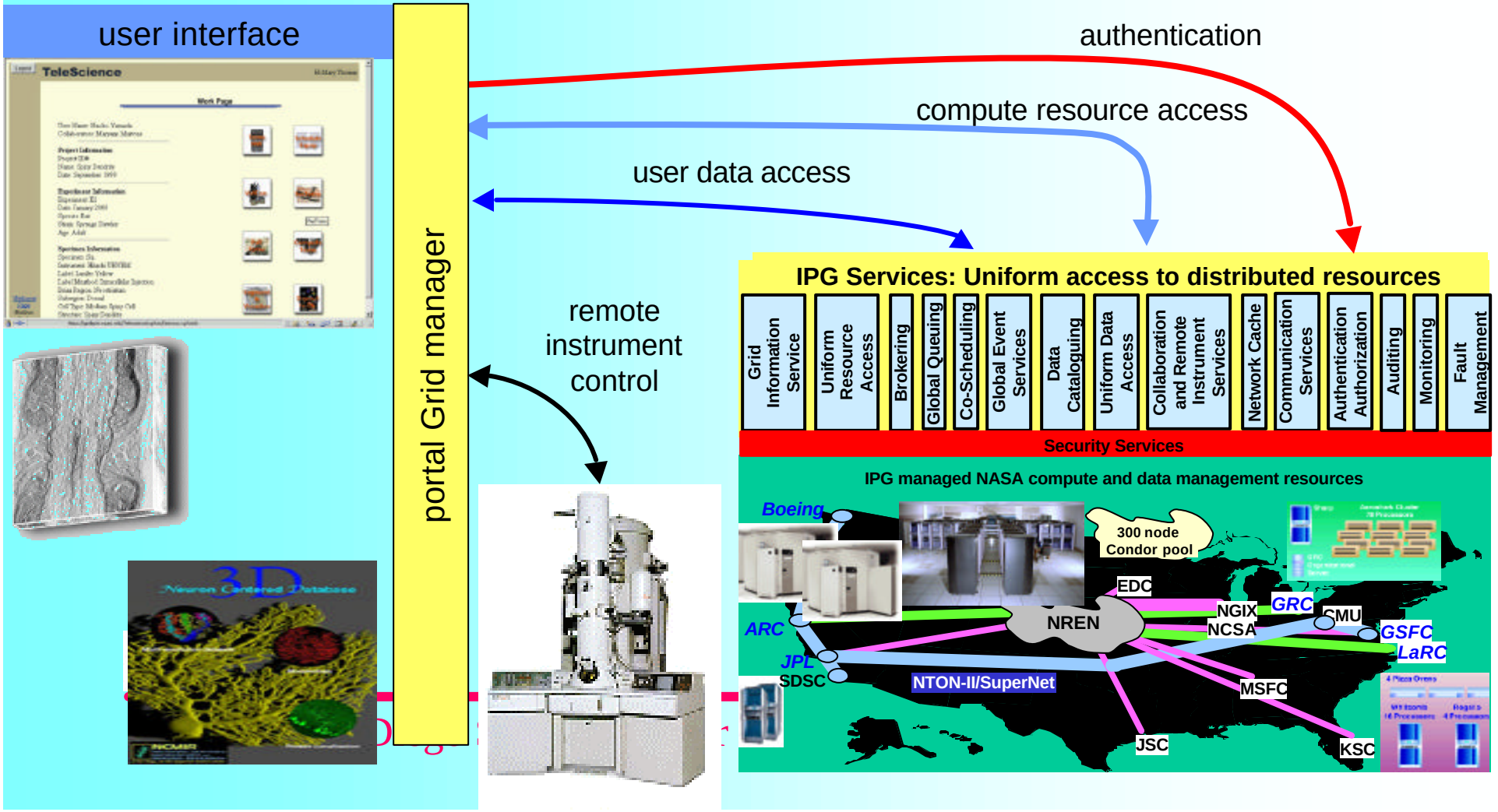

Clients
Web Browser
http, https
Grid App. Portals
Discipline / Application Specific Portals (e.g. SDSC TeleScience)
Problem Solving Environments (AVS, SciRun, Cactus)
Environment Management (LaunchPad, HotPage)
composition frameworks
Python, Java, Perl, etc., JSPs format to html
Apache SOAP, .NET, etc.
XML / SOAP over Grid Security Infrastructure
Grid Web Services
Job Submission / Control
File Transfer
Data Management
Monitoring
Events
.....
Credential Management
Workflow Management
other services:
•visualization
•interface builders
•collaboration tools
•numerical grid generators
•etc.
CoG Kits implementing Web Services in servelets, servers, etc.
Apache Tomcat&WebSphere &Cold Fusion=JVM + servlet instantiation + routing
Grid Protocols and Grid Security Infrastructure
Grid Services Collective and Resource Access
CORBA
GRAM
Condor-G
SRB/MCAT
Replica Catalog / Management
GridFTP
Grid Monitoring Architecture
Grid X.509 Certification Authority
MPI
Grid Information Service
Secure, Reliable Group Comm.
Grid Web Service Description (WSDL) & Discovery (UDDI)
Grid Protocols and Grid Security Infrastructure
Resources
Compute (many)
Storage (many)
Communi-cation
Instruments (various)

NASA/IPG 4Q01 Milestone:Telescience Portal

- **Telescience portal met NASA/IPG 4Q01 Milestone:**
  - **Remote access to high data-rate instruments**
- **Description:**
  - **Demonstrate using Grid services for remote connectivity to high data-rate instruments and distributed real-time access to instrument data.**
- **Output:**
  - **One or more instruments; data rate exceeding 50 Mbits/sec**
- **Outcome:**
  - **Use of Grid services and infrastructure for on-demand connectivity to high data-rate instruments and enhanced engineering and scientific collaboration**
    - **Access SRB Collections on IPG resources**
    - **Using GSI for security, Globus GRAM for related tasks**
    - **Portal based on GridPort Toolkit**

Remote access to high data-rate instruments

The UC San Diego Telescience for Advanced Tomography



NWS activities, R.Wolski UCSB

- Performance tuning
  - Internal caching
  - LDAP wire protocol benchmarked as a replacement for internal NWS messaging protocol (NWS is faster)
  - Order of magnitude improvement in global fetch speed (i.e. fetch *.*)
- GIS and GMA motivated restructuring ala GGF
  - The NWS conforms to the GMA
  - New NWS caching Index Node for the GIS
    - **GIS conformance**
    - **Further performance enhancement**

NWS activities for next year

- **Packaging and Configuration**
  - **NWS will use the MDS-2 as a name service**
  - **Globus packaging and deployment (GPT)**
  - **One-button install and configure, NMI release 1.**
- **NWS-DB**
  - **Relational database archive of performance data**
  - **Prototype at SC01**
- **New sensor registration and integration interfaces**
  - **Improves extensibility**
  - **Thread-safe messaging**
  - **Better, cleaner, cheaper, less-filling, …**
- **Long-range forecasting**

APST - Scheduling Research, H.Casanova

UCSD (casanova@cs.ucsd.edu)

- Goal: provide scheduling methodology for parameter sweep applications (APST provides scheduling capabilities beyond those of the NASA ILAB project.)
- Progress: extended the scheduling methodology
  - Better adaptation to poor resource/application information
  - Scheduling with mixed resources (batch, interactive, condor)
  - Improved task duplication strategy
- Ongoing work: modularizing the scheduling methodology
  - Reusable externally as a *scheduling service*
  - APST scheduling can be used w/o using APST deployment
  - Important for impact on NASA work (ILAB)

APST - software, H.Casanova UCSD

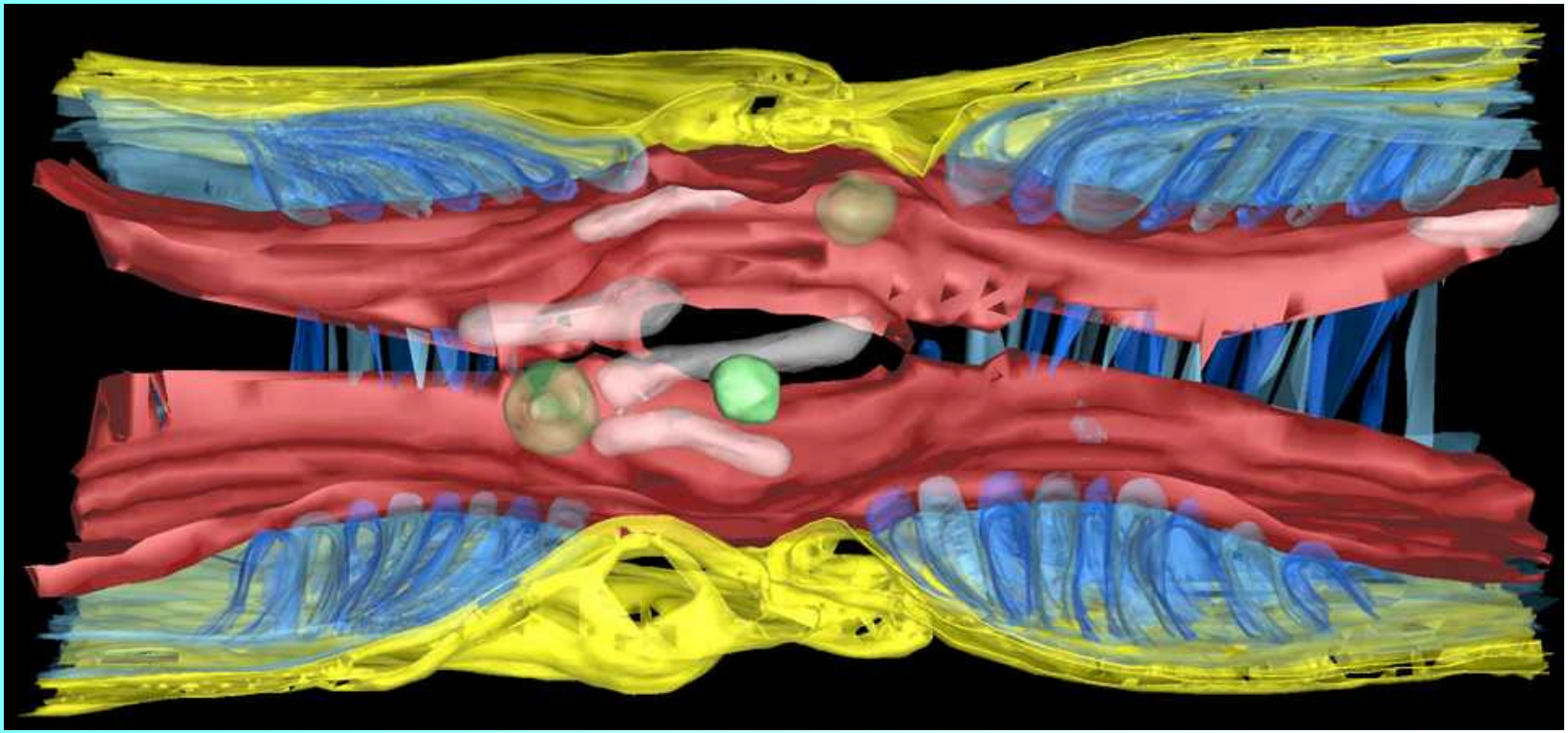
- APST v1.1 to be released before Xmas
- New functionalities
  - Supports Condor resources
  - Prototype GridFTP support
  - Better supports use of batch resources
  - Supports “light” task dependencies
- Increased Usability
  - Hierarchical XML resource description
  - Supports user-directed data staging
  - Software re-engineered and hardened
  - Tested with over 10 applications in the last year

Telescience and Grid

- Remote access to high performance instruments and computational resources available across domains and platforms from the web
- Grid computing for the non-computer scientist
  - Point and click interface to grid computing
- Fully automated retrieval of raw data and depositing of processed data
- Seamless integration with SRB

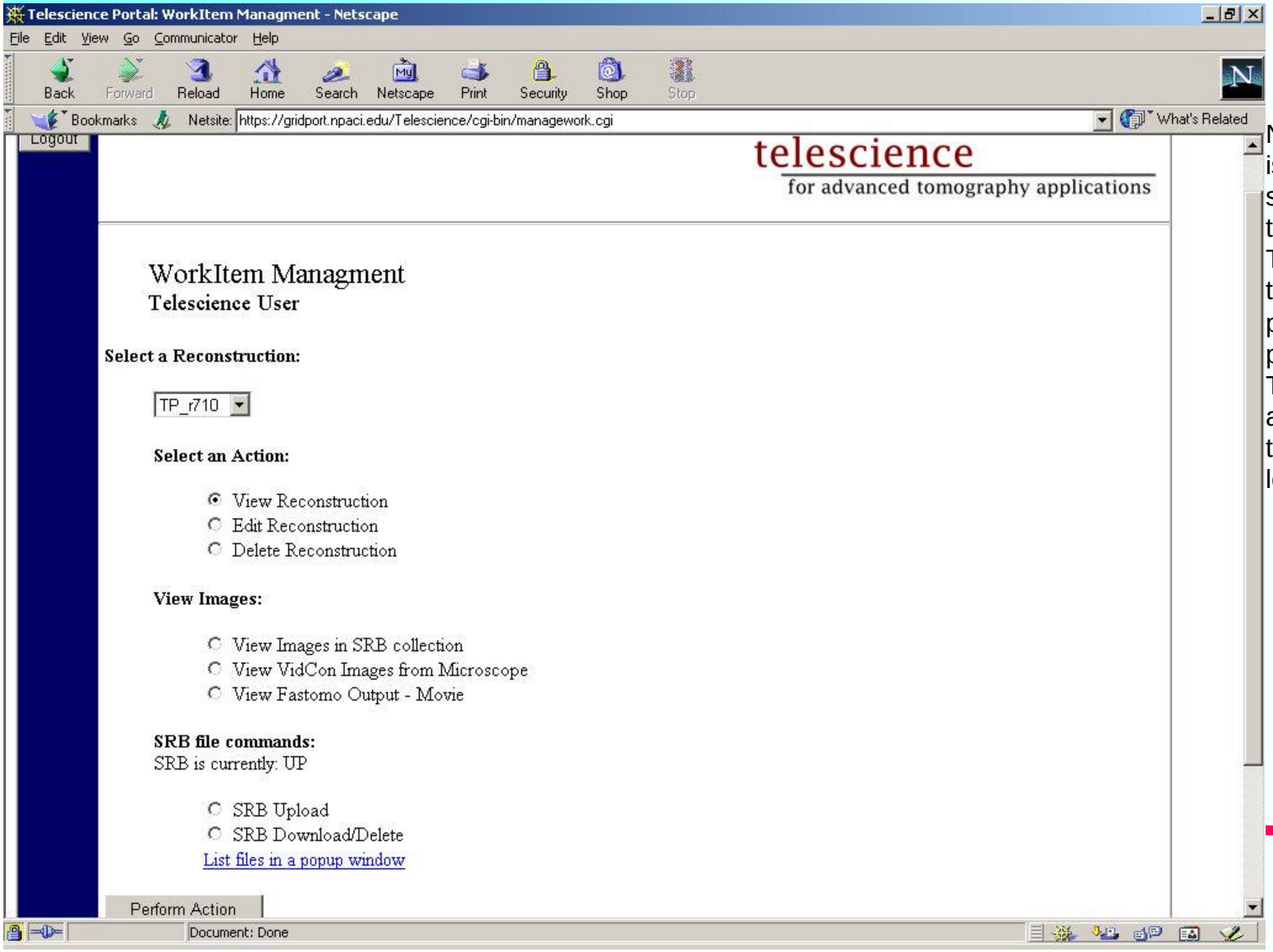
Telescience Example

- Example of a fully reconstructed Node of Ranvier



Telescience Portal SDSC - Manage work page using SRB:

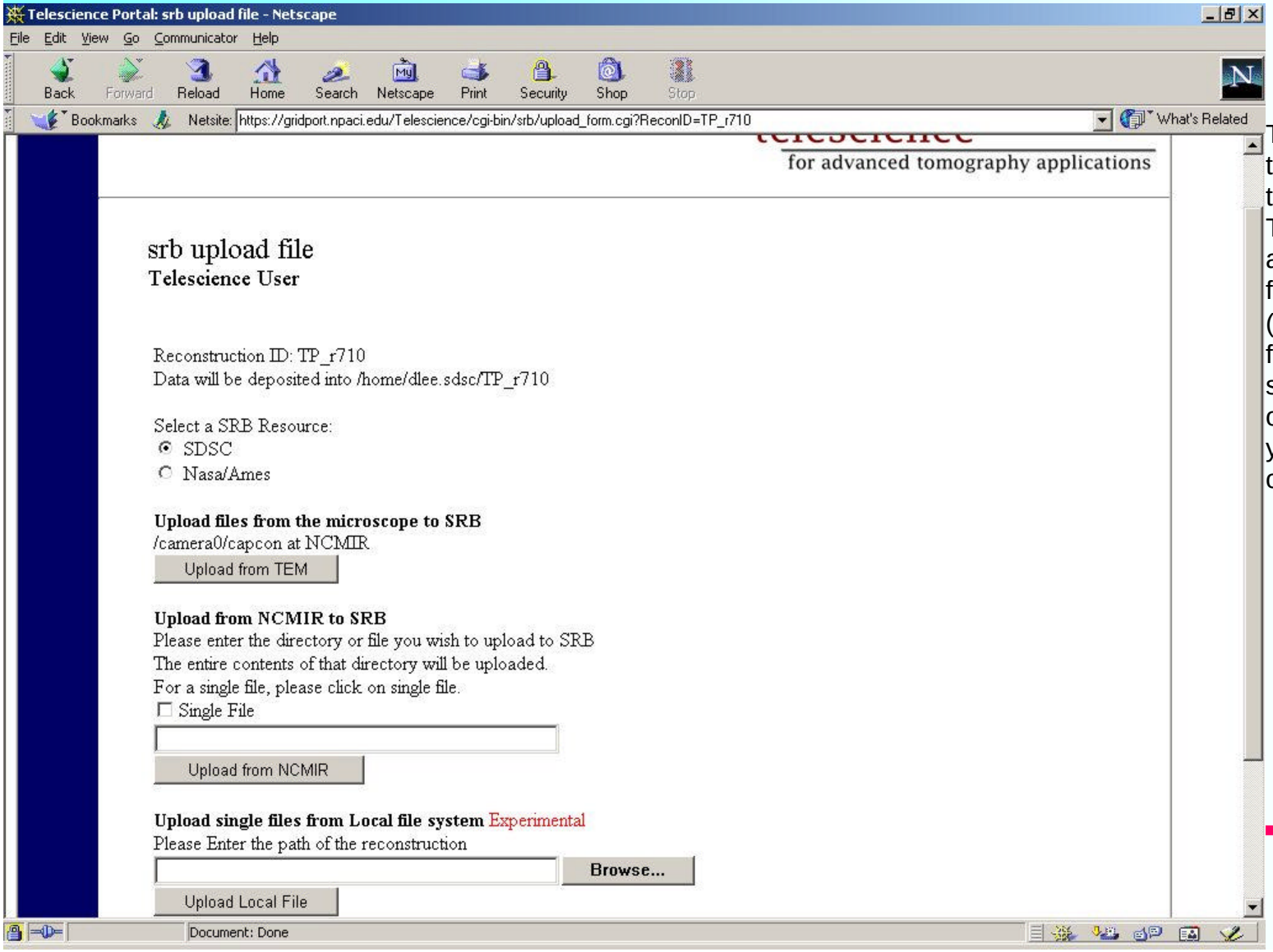
From this page the user can view/edit/delete reconstructions as well as view images, upload, and download files from SRB.



Near the top of the page, is an area where you can select the reconstruction to be modified or viewed. The user then selects the task which they wish to perform and press the perform action button. The radio buttons are approximately ordered in the highest to lowest level function.

Telescience Portal SDSC- SRB data upload:

This page is where the user can move the data from the microscope to the appropriate SRB collection. The user simply selects the resource they would like to use and press upload. More advanced upload functions are also available for use.



There are several ways to move your data into the portal and SRB. These methods accommodate moving files from the microscope (first option), moving data from the NCMIR file system (second option), or uploading data from your local computer (third option).

Porting OpenLDAP and Globus to the Cray T90 and SV1

- The problem: Globus depends on OpenLDAP; OpenLDAP depends on the existence of a 32-bit integer type; the T90 and SV1 have no such type.
- The good news: Only a small part of OpenLDAP actually depends on this, and Globus doesn't use that part

Cray T90 and SV1 port (cont.)

- Globus 1.1.4+ is now running on SDSC's T90 and Utexas's SV1. It's ready to be deployed on NASA's SV1. A Tarball with the installation has been provided.
- The modifications have not yet been checked back into the mainstream OpenLDAP or Globus sources. Patched versions are available (contact kst@sdsc.edu).

GSI-FTP-SRB

- GSI/FTP server front-end
- SRB back-end

GridFTP/SRB Integration

- The goal: Make SRB space visible to GridFTP.
- Progress so far: A working preliminary implementation uses the “transparency” feature of SRB. This intercepts system calls so that SRB space appears as part of the Unix file system; gsiftpd runs on top of this.
- This effectively tests compatibility of the GSI authentication system implementations in Globus and SRB

GridFTP/SRB Integration (continued)

- As far as the gsiftpd daemon is concerned, SRB space is just a /srb directory tree.
- To do:
  - Modify gsiftpd so it makes calls to the SRB API directly.
  - Evaluate performance of the GSI authentication

Future Projects

- Web Services Development
- Discipline Specific Portals
- Discipline Specific Collections (Darwin)
- Distributed Data Collection between NASA Ames and Goddard (DAO)
- Automated Collection Management creation for production applications